

# Research on Signal Processing Architecture Design and Compatibility of Multimodal Perception System for Emergency Rescue Unmanned Aerial Vehicles

Jiahang Zhu

Anhui University of Industry, Ma'anshan, Anhui, China

**Copyright:** © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

**Abstract:** In response to the core pain points in emergency rescue scenarios, such as the strong heterogeneity of multi-modal sensing signals, low efficiency of collaborative processing, and insufficient hardware adaptability, this paper reviews the application status of multi-modal sensing technology in emergency rescue fields, analyzes the key issues of the existing signal processing system, and constructs a three-level modular signal processing architecture of “sensing–transmission–pre-processing”. The sensing layer inputs the raw signals and outputs signals in a unified data format; the transmission layer inputs multiple signals, outputs these signals with priority, different paths, and strategies; the pre-processing layer processes the input signals and transmits the information to the unmanned aerial vehicle (UAV) terminal. This research clearly defines the core principles and modular logic of the architecture design, and explores the transmission optimization, resource allocation, and hardware compatibility schemes of different signal types from a technical perspective, demonstrating the balance between the general and specific paths of the architecture. The research results can provide architecture-level design references for the multi-modal sensing module of the airborne-ground integrated emergency rescue UAV, solve the problem of heterogeneous signal collaborative processing, and improve the overall response efficiency and reliability of the emergency rescue system.

**Keywords:** Emergency rescue unmanned aircraft; Multi-modal perception; Signal processing architecture; Modular design; Technical adaptability

**Online publication:** May 21, 2026

## 1. Introduction

### 1.1. Research background

In recent years, sudden natural disasters such as earthquakes and forest fires have occurred frequently. The rapid response within the “golden 72 hours” is the key to reducing the casualty rate <sup>[1]</sup>. With the frequent occurrence of extreme disaster accidents, emergency rescue scenarios present complex, variable, high-risk

and time-sensitive characteristics. Traditional manual rescue and single-sensor unmanned aerial vehicles (UAVs) have been unable to meet the demands for all-encompassing, rapid and precise rescue. With the rapid development of low-altitude economy and intelligent equipment technology, UAVs have become the core carrier of emergency rescue. However, the disaster-stricken environment is often extremely complex: the thick smoke at the fire scene blinds visible light cameras, the low temperature at night and in the ruins deepens, causing the false alarm rate of single infrared detection to soar, and the high packet loss rate caused by the damaged communication infrastructure during the disaster makes the perception system of UAVs face severe challenges. Therefore, multi-modal perception technology is the key support for achieving precise perception and intelligent decision-making of UAVs, and can cover multiple scenarios such as night rescue, life sign detection, and environmental situation perception.

The advancement of deep learning algorithms has significantly enhanced the target recognition accuracy of multimodal perception systems. For instance, on the public drone perspective target detection dataset (such as VisDrone), the average precision (mAP) of mainstream algorithms has increased from approximately 40% of the early models to over 70% of the current advanced models, providing technical possibilities for the precise implementation of emergency rescue <sup>[2]</sup>. However, traditional research on multimodal fusion only focuses on “target detection accuracy (mAP)” while ignoring the limitations of “resource constraints”, resulting in the following three major problems for emergency rescue drones: they generally rely on a single sensor or simple image overlay, lack a deep multi-modal complementary mechanism, and have difficulty maintaining continuous perception capabilities in extreme environments; the collection protocols, transmission rates, and data formats of different types of perception signals vary significantly, leading to low efficiency in signal collaborative processing; emergency rescue sites often require the urgent dispatch of multi-brand, heterogeneous drone clusters, while existing systems are mostly tightly coupled in software and hardware, lacking universal adaptation interfaces, resulting in a serious “data island” phenomenon among devices. In this context, exploring a general architecture for multi-modal signal processing of UAVs in emergency rescue scenarios with high robustness and resource adaptability has become the core issue for improving the intelligence level of emergency rescue UAVs.

## **1.2. Research significance**

### **1.2.1. Theoretical significance**

This research focuses on the heterogeneous characteristics of multimodal sensing signals in emergency rescue scenarios, and constructs a three-level signal processing architecture of perception–transmission–preprocessing. It breaks through the limitations of traditional tightly coupled design and proposes standardized interfaces and hierarchical processing models. It solves the coordination barriers caused by differences in data volume of heterogeneous signals (infrared/visible light/positioning), providing a new method for system modeling in complex scenarios and enriching the theoretical system of multi-modal sensing technology of UAV in the field of emergency rescue.

This research also demonstrates the design logic and adaptation path of the modular architecture from the technical perspective, providing a theoretical framework for the standardization and modularization research of UAV signal processing architectures, and promoting technological theoretical innovation in the field of emergency rescue intelligent equipment. The multimodal large model, as the core support for complex environment interaction of embodied intelligent unmanned systems, its modal alignment and feature fusion technology provides a new theoretical perspective for heterogeneous signal processing. The

architecture design in this paper can further enrich the application theory in this field<sup>[3]</sup>.

### **1.2.2. Practical significance**

The signal processing architecture proposed in this study can effectively address the practical problems of low efficiency and poor hardware compatibility in the collaborative processing of multi-modal signals of emergency rescue drones, and reduce the integration cost and debugging difficulty of different signal modules. At the same time, the signal processing architecture design in this paper has two application advantages as follows:

- (1) The modular design of the architecture can quickly adapt to different types of drone hardware platforms and rescue scene requirements, improve the real-time performance of signal transmission and the overall reliability of the system, provide a feasible architecture reference for the engineering application of air-ground integrated emergency rescue drones, and help improve the response speed and rescue success rate of emergency rescue;
- (2) The signal processing architecture, with standardized interfaces and cross-architecture adaptation algorithm libraries, supports plug-and-play of sensors from different brands and flight platforms, compresses the system integration and debugging time from hourly level to minute level, significantly reduces the integration threshold and cost of rescue equipment. In complex rescue scenarios such as urban underground spaces, this architecture can effectively solve the problems of difficult disaster information acquisition and blocked emergency communication, providing technical support for special environment rescue<sup>[4]</sup>.

## **2. Current status of multi-modal perception signal processing for emergency rescue UAVs**

### **2.1. The current application status of multimodal perception technology**

Currently, multimodal perception technology has been gradually applied in the field of emergency rescue drones. Infrared sensing signals can achieve nocturnal life sign detection and fire source location, while visible light sensing signals are used for environmental situation visualization and personnel identification. Positioning signals provide precise spatial coordinates for the drone and the rescue target. The multimodal perception technology of emergency rescue drones has initially achieved a multi-dimensional expansion from a single visible light mode to “infrared + visible light + positioning + laser radar”. However, commercial payloads such as DJI Zenmuse H20T have achieved four-modal integration at the hardware level and demonstrated certain efficacy in fire source location and nocturnal search and rescue. However, their technical essence still belongs to the physical stacking of sensors. Existing applications generally have a tendency of “heavy collection, light processing”. Most products focus on the integration of sensor hardware and lack systematic planning for the signal processing process. The signal processing is mostly “serial” designed, and no unified modular architecture has been formed, resulting in difficulties in achieving efficient coordination of different sensor signals.

### **2.2. The current application status of multimodal perception technology**

The current emergency rescue UAV signal processing architecture mainly falls into two categories: one is a single signal processing architecture, which is designed only for infrared or visible light signals, with a

simple processing flow, but it cannot meet the multi-modal fusion rescue requirements; the other is a simple fusion processing architecture, which realizes the splicing of multiple signals through a simple interface, lacking the adaptation to signal heterogeneity, and has problems such as high transmission delay and large resource occupation <sup>[5]</sup>.

In addition, the mainstream architectures mostly adopt a “software and hardware tightly coupled” design, with signal acquisition, transmission, and processing logic being fixed on a specific hardware platform. Once the sensor model or UAV platform (such as switching from ARM architecture to X86) is changed, it often requires reconfiguration of the underlying code, lacking standardized interfaces and module decoupling mechanisms, resulting in a long system integration cycle and high maintenance costs. At the same time, the existing architectures do not fully consider the dynamics of emergency rescue scenarios, have insufficient compatibility design for different hardware platforms, and are difficult to adapt to the hardware characteristics of mainstream unmanned aerial vehicle models, limiting the large-scale application of the technology.

### **2.3. Technical support and research foundation**

At the hardware level, the rapid development of the low-altitude economy has provided a vast application space for UAV technology. Currently, mainstream UAV platforms have gradually acquired multi-sensor expansion interfaces, enabling the simultaneous deployment of various sensing devices such as visible light cameras and thermal infrared cameras, providing hardware support for multimodal perception systems <sup>[6]</sup>.

At the signal processing level, edge computing shifts computing tasks from the cloud to the data source vicinity, significantly reducing signal transmission latency. In the UAV-assisted edge computing system, the UAV can not only collect sensing data but also complete preprocessing tasks locally. This technical approach effectively resolves the contradiction between real-time processing of multimodal signals and the limited computing power of the UAV, providing technical support for the implementation of the “perception-transmission-preprocessing” three-level architecture in this paper <sup>[7]</sup>.

At the same time, the actual needs of existing emergency rescue scenarios have also provided clear application guidance for the architecture design, laying a solid foundation for the technical demonstration and architecture design of this paper. The breakthrough in the application of network coding technology in emergency communication has significantly improved the broadcast throughput of UAV relays, providing new technical support for the efficient transmission of multimodal signals <sup>[8]</sup>.

## **3. The core issue of multi-modal perception signal processing for emergency rescue unmanned aircraft**

### **3.1. The current application status of multimodal perception technology**

Infrared, visible light, and positioning signals have significant differences in data characteristics, transmission protocols, and processing requirements, forming a barrier of heterogeneity. This is the primary challenge in current multimodal perception signal processing. From the perspective of data characteristics, the parameter differences of different modal signals are extremely significant: infrared signals typically use a 16-bit depth grayscale image format, with a sampling frequency ranging from 15 to 30 FPS, and the single-frame data volume is approximately 1.2MB; visible light signals are mostly in the 24-bit RGB format, with a resolution of up to 4K in high-definition mode, and the single-frame data volume is as high as 12MB, with a sampling

frequency of 30–60 FPS required to ensure video smoothness; positioning signals are discrete coordinate data, with a text format, and the sampling frequency is only 10–20 Hz, and the single data volume is less than 1KB. This huge difference in data volume easily leads to extreme situations such as data congestion or resource idleness in the same processing channel <sup>[9]</sup>.

At the transmission protocol level, the adaptability difficulty brought by heterogeneity is also prominent. Infrared and visible light signals adopt mostly Ethernet or high-speed wireless transmission protocols to meet the demand for large bandwidth transmission; while positioning signals use mostly serial ports or low-power Bluetooth for transmission, with a lower transmission rate but high stability requirements. The interface conversion between different protocols not only increases hardware complexity but also introduces additional transmission delays. For example, the conversion delay from serial port to Ethernet can reach 50–100ms, which is not negligible in the real-time requirements of emergency rescue <sup>[10]</sup>.

More importantly, the acquisition timing of different signals is difficult to be unified, which makes data synchronization challenging. For instance, the exposure time of the visible light camera differs from the detection cycle of the infrared sensor. If precise timing calibration is not carried out, there will be a problem of target position deviation. In the scenario of life sign detection, this may lead to misjudgment of the survivor's position, seriously affecting the rescue efficiency. Moreover, there is a conflict between the processing accuracy and real-time requirements of different modal signals: when infrared signals are used for life sign detection, the temperature recognition accuracy must be within  $\pm 0.5^{\circ}\text{C}$ . Otherwise, environmental heat sources may be mistakenly identified as survivors; when visible light signals are used for target recognition, the feature extraction of a single frame image must be completed within 300ms to meet the real-time tracking requirements during the unmanned aircraft's dynamic flight; and the positioning signal must ensure that the positioning error does not exceed 1m; or it will affect the precise hovering rescue operation of the unmanned aircraft <sup>[11]</sup>.

This multi-dimensional heterogeneity and differences accumulate, forming a technical barrier for collaborative processing. The existing architecture lacks a unified heterogeneous adaptation mechanism, making it difficult to achieve efficient fusion of multimodal signals.

### **3.2. Problem of missing modularization in signal processing architecture**

The current signal processing flow lacks a unified modular design, with unclear division of functional modules, and the perception, transmission, and pre-processing stages have not formed standardized interfaces. The coupling degree between modules is high, which leads to the fact that adjustments to one module will affect the operational stability of the entire system. Research shows that modular payload configuration is a key design concept for enhancing the scalability of UAV systems, but the existing architecture lacks this design, resulting in insufficient system adaptability <sup>[12]</sup>.

Specifically, sensors and processing algorithms are usually hard-coded into specific hardware platforms, lacking standardized abstract interface layers. Once the sensor model needs to be changed (such as upgrading from single-light to dual-light) or the UAV carrier is switched, it often involves the re-development of the underlying driver and the full-system integration, resulting in a long equipment upgrade cycle and high costs, and making it impossible to achieve “plug-and-play” agile deployment. For example, when replacing an infrared sensor, the code logic of the transmission layer and pre-processing layer needs to be modified simultaneously, resulting in a long debugging period and high costs, making it difficult to meet the demand for rapid equipment replacement in emergency rescue scenarios. Secondly, the architecture lacks

generalizability and cannot adapt to the hardware configuration differences of different types of UAVs. The computing units, communication modules, and power supply systems of mainstream UAV platforms have significant differences, and the existing architecture has not established a universal adaptation standard, making it difficult for the same signal processing system to run stably on different platforms. It is worth noting that software-defined radio (SDR) technology enables flexible adjustment of signal processing strategies through software, allowing for rapid switching between different tasks, providing a design concept for modular interface design<sup>[13]</sup>.

Meanwhile, the architecture has weak adaptability for specific scenarios, making it difficult to adjust the signal processing strategies according to different rescue scenarios. In the fire rescue scenario, the detection accuracy and anti-interference ability of infrared signals need to be improved first. In the earthquake search and rescue scenario, the real-time performance of positioning signals and the target recognition ability of visible light signals need to be enhanced. In the flood rescue scenario, the anti-interference and stability of signal transmission need to be optimized. The existing architecture lacks a scenario-specific parameter configuration mechanism, and it is unable to dynamically adjust the acquisition parameters, transmission strategies, and processing algorithms according to the requirements of the scenarios, resulting in extremely poor flexibility.

Furthermore, the lack of modularity also makes system upgrades difficult. The access of new sensors requires the reconfiguration of the entire signal processing flow, and it cannot be achieved through plug-and-play, which restricts the iterative upgrade of multi-modal perception technology.

### **3.3. Hardware compatibility and resource optimization issues**

The issues of hardware compatibility and resource optimization both stem from the fact that static configurations cannot meet dynamic demands.

In terms of hardware compatibility, the mainstream emergency rescue drone hardware platforms have architectural differences. The existing signal processing architecture does not fully consider hardware adaptability, resulting in high difficulty and poor compatibility in integrating sensor modules with the drone platform. From the perspective of computing units, consumer-grade drones mostly adopt ARM architecture processors, while industrial-grade drones commonly use X86 architecture or FPGA. There are differences in instruction sets, computing power allocation, and interface standards among different architectures. The existing architecture has not designed cross-architecture adaptation solutions, leading to significant differences in the running efficiency of signal processing algorithms on different platforms. Additionally, some algorithms are not even compatible. From the perspective of communication modules, the communication methods of drones include WiFi, 4G/5G, data transmission radios, and satellite communication, and different communication methods have different bandwidths, delays, and anti-interference capabilities. The existing architecture has not established a dynamic communication adaptation mechanism, and in complex rescue scenarios, it is prone to signal interruption and transmission delays.

In terms of resource optimization, the hardware resources of emergency rescue drones are limited, and multi-modal signal processing requires high resource consumption. The existing architecture has not established an effective resource allocation mechanism. High-bandwidth visible light signals easily occupy a large amount of computing power and storage resources, resulting in increased processing delays for infrared and positioning signals, and even data loss. At the same time, the signal processing algorithms are mostly designed in a general way and have not been lightweight-optimized for the characteristics of drone hardware.

The high computing power occupation rate shortens the endurance time of the drone, affecting the duration of rescue operations. For this problem, existing research has achieved a real-time inference speed of 650 frames per second on the Jetson Orin Nano platform, proving that through model compression and hardware optimization, efficient signal processing can be completed under limited resources<sup>[14]</sup>. It can be seen that lightweight design is feasible. In addition, the resource scheduling lacks a priority mechanism and does not set core signals such as vital sign detection and emergency positioning as high priorities. In resource-scarce situations, it is prone to lag in the processing of core information, unable to meet the “every second counts” time requirements of emergency rescue.

### **3.4. The problem of insufficient signal fusion and collaborative decision-making capabilities**

The core value of multimodal signals lies in their complementary integration. However, existing systems generally suffer from insufficient integration depth and weak collaborative decision-making capabilities. From the perspective of integration levels, most systems only achieve simple data-level concatenation and do not conduct deep integration at the feature and decision-making levels, thus failing to fully leverage the advantages of different modal signals. For instance, merely superimposing and displaying infrared images and visible light images without extracting their features for cross-validation makes them vulnerable to environmental interference and prone to misjudgment; without spatially fusing the positioning signal and the image signal, it is impossible to precisely mark the target location, thereby reducing rescue efficiency.

Furthermore, academic research indicates that single-modal data provides limited information in complex scenarios, and existing fusion methods mostly adopt simple concatenation or weighted summation, lacking modeling of the complex interaction relationships between modalities, and facing significant challenges in feature alignment, semantic consistency, and other aspects<sup>[15]</sup>.

The insufficient collaborative decision-making ability is manifested in the lack of a unified decision-making logic and feedback mechanism. The data processed by multimodal signal processing has not formed standardized decision inputs, and the unmanned aircraft flight control and rescue instructions issuance cannot be achieved through the fusion results to realize autonomous collaboration. For example, after infrared detection of a heat source, it cannot automatically linkage the positioning signal to lock the position and adjust the unmanned aircraft’s flight attitude, and still requires manual intervention, resulting in a slow response speed. At the same time, the system lacks a real-time feedback mechanism, and cannot dynamically adjust the acquisition parameters and transmission strategies based on the signal processing results, resulting in poor adaptability in complex environments. Moreover, in multi-unmanned aircraft collaborative scenarios, the multimodal signal processing architectures of each unmanned aircraft operate independently, lacking data sharing and collaborative decision-making interfaces, unable to form an integrated rescue perception network between air and ground, which restricts the overall rescue efficiency improvement.

According to the technical forum discussion held by the Chinese Computer Society in November 2025, the current AI visual models of unmanned aircraft in emergency rescue applications have core bottlenecks such as weak multimodal interaction ability, large gap between practical and simulation, difficulty in obtaining disaster samples, and insufficient generalization ability of the models. Experts pointed out that the existing systems are difficult to cope with appearance changes such as target deformation and color change, and the complexity of actual scenarios far exceeds the processing capacity of the models. More critically,

the task requirements are shifting from simple target perception to overall situation perception in complex environments, requiring models to have strong autonomous collaboration and adaptive capabilities, but the existing technologies are unable to meet this transformation.

## **4. Design and adaptation countermeasures for multi-modal perception signal processing architecture of emergency rescue UAVs**

### **4.1. Core principles and objectives of architecture design**

#### **4.1.1. Core principles**

The modularization principle involves breaking down the signal processing process into independent and scalable functional modules, reducing the coupling between modules and enabling plug-and-play functionality and independent upgrades for the modules. This signal processing architecture is regarded as a small embedded system, with each module working flexibly together. The benefits of this approach are that the system is highly operable and it is easy to locate the faulty module when there is a failure.

The balance between generality and specificity principle involves two aspects: the first is to make the basic framework of the signal processing architecture adaptable to multiple types of UAV platforms, allowing different types of UAVs to be applicable to this system; the second is to reserve scenario-specific adaptation interfaces to meet the specific needs of different rescue scenarios. For different rescue scenarios, different specific modules are connected to meet the actual situation. This system should balance the relationship between generality and specificity and propose feasible solutions.

The priority principle of real-time performance and reliability involves optimizing the efficiency of signal transmission and processing to ensure the timeliness requirements of emergency rescue scenarios, reducing the total rescue time to within the controllable “golden 72 hours” to meet the speed requirements of actual rescue<sup>[1]</sup>. At the same time, improving the system’s anti-interference ability and stability, achieving both real-time performance and reliability within the required standards, and balancing the trade-offs between these two factors.

The principle of efficient resource utilization involves designing lightweight algorithms and dynamic resource allocation mechanisms to achieve the optimal configuration of computing power, power, and storage for the UAV.

#### **4.1.2. Design objective**

The system aims to achieve the goals of unified acquisition, collaborative transmission, and efficient preprocessing of multimodal signals. The solution strategy is to build a three-level architecture of perception–transmission–preprocessing, and break down this project into three steps, gradually simplifying the complex tasks.

The system aims to achieve the goal of enhancing the hardware adaptability and scene flexibility of the architecture. The solution strategy is to clarify the interface standards and coordination mechanisms of each module, determining the functions and connection methods of the modules like an embedded system, so as to facilitate the replacement of modules in the future.

The system aims to achieve the goal of stable and efficient processing of multimodal sensing signals in emergency rescue scenarios. The solution strategy is to reduce the resource occupation and transmission delay of signal processing, and compress these indicators from all aspects.

The system aims to support autonomous rescue operations of unmanned aircraft. The solution strategy is to deeply integrate the signal feature layer and the decision-making layer to enhance its collaborative decision-making ability.

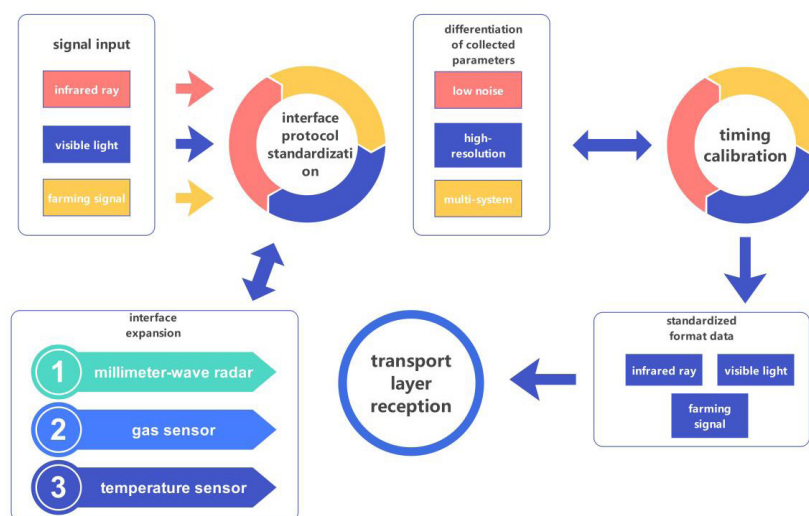
## 4.2. The construction of a three-level architecture consisting of perception, transmission, and pre-processing

### 4.2.1. Perception layer: Standardized multi-modal signal acquisition module

The perception layer serves as the signal input entry point. It designs standardized sensor interface protocols to unify the acquisition timing and data format for infrared, visible light, and positioning signals, enabling plug-and-play functionality for different types of sensors. According to the characteristics of different signals, differentiated acquisition parameters are configured: infrared signals are optimized with low-noise acquisition parameters, using non-cooled infrared detectors to control the temperature recognition accuracy within  $\pm 0.3^{\circ}\text{C}$ , thereby improving the accuracy of life sign detection at night; visible light signals are adapted to high-resolution acquisition requirements, supporting 4K/1080P resolution switching, and dynamically adjusting the sampling frequency based on the scene to ensure image quality while reducing data volume; positioning signals integrate GPS, Beidou, and GLONASS multi-systems, using RTK technology to reduce positioning errors to within 0.1m, ensuring precise synchronization of spatial coordinates.

A unified timing calibration mechanism is established, using hardware synchronous clocks and software timestamps for calibration to achieve precise synchronization of different sensor signals and eliminate target position offset caused by timing deviations. At the same time, the perception layer reserves expansion interfaces to support the access of new types of sensors such as millimeter-wave radars, gas sensors, and temperature and humidity sensors, meeting the diverse perception needs of complex rescue scenarios. The perception layer and the transmission layer adopt standardized Ethernet interfaces to reduce interface conversion delays and improve signal transmission efficiency.

The processing flow of the perception layer is shown in **Figure 1**.



**Figure 1.** The processing flow of the perception layer.

#### **4.2.2. Transport layer: Heterogeneous signal co-transmission mechanism**

The transport layer designs a hierarchical transmission strategy to address the heterogeneous characteristics of multi-modal signals and resolve conflicts in the transmission requirements of different signals. High-bandwidth signals use WiFi 6 or 5G high-speed wireless communication protocols, configure large bandwidth transmission channels, and control the transmission delay within 50ms; low-bandwidth signals use low-power Bluetooth or data transmission radios for transmission, optimize resource occupation, and ensure transmission stability. A signal priority scheduling mechanism is established, setting vital sign detection and emergency positioning signals as high priority, and prioritizing the allocation of transmission resources; environmental perception and status monitoring signals are set as low priority, transmitting when resources are sufficient, ensuring that core information reaches first.

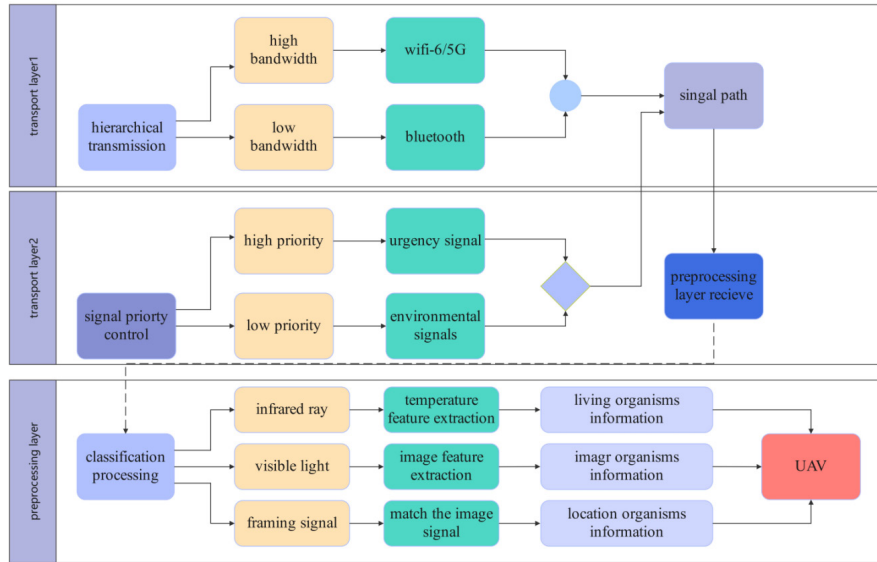
Network coding technology is introduced to enhance the reliability and throughput of emergency communication. In complex rescue scenarios, through relay transmission by drones, network coding technology is used to reduce the number of data retransmissions, improve the anti-interference ability of signal transmission, and solve the problem of blocked emergency communication. At the same time, the transport layer designs a dynamic bandwidth allocation processing strategy, adjusting the transmission bandwidth of each signal in real time according to signal type, scene requirements and channel quality, avoiding data congestion and resource idleness, and achieving collaborative and efficient transmission of heterogeneous signals.

#### **4.2.3. Preprocessing layer: Classification-based signal processing module**

The preprocessing layer is responsible for cleaning, integrating and preliminarily processing the signals, designing classification processing logic, and improving processing efficiency and accuracy. For infrared signals, an adaptive filtering algorithm is used to remove environmental noise, extract temperature features and vital sign features, and identify survivors and fire source locations. For visible light signals, image denoising, distortion correction, target detection and tracking are completed. A lightweight deep learning model is adopted, and the feature extraction of a single frame image is completed within 300ms to meet real-time requirements. For positioning signals, coordinate calibration, error correction and spatial fusion are carried out, and the positioning results are matched with the image signals to precisely mark the target location.

Edge computing technology is used to implement the preprocessing function, and the signal processing tasks are deployed on the local computing unit of the UAV to reduce the pressure of data transmission to the backend and lower latency. The preprocessing layer designs modular processing units, and the processing units for each signal operate independently without interfering with each other. At the same time, a fusion interface is reserved to achieve cross-validation of the feature layer. For example, the thermal source features detected by infrared detection and the human features recognized by visible light are fused to eliminate environmental heat source interference and improve the accuracy of target recognition; the positioning signal and the image signal are fused to generate a rescue map with coordinates, providing precise support for rescue decisions.

The processing methods of the transport layer and the pre-processing layer are shown in **Figure 2**.



**Figure 2.** The processing methods of the transport layer and the pre-processing layer.

### 4.3. Architecture adaptability optimization strategy

#### 4.3.1. Hardware compatibility and adaptation

This paper analyzes the characteristics of the mainstream emergency rescue UAV hardware platforms, designs a universal hardware interface for this system, clarifies the connection standards between the architecture and the UAV flight controller, computing unit, and communication module, and realizes the rapid adaptation of the architecture to different hardware platforms. In addition, this is also the advantage of this system in terms of hardware adaptation.

This system also reserves hardware expansion interfaces to support the connection of new sensors, computing modules, and communication modules, enhancing the scalability of the architecture. Establishes hardware compatibility testing standards, conducts compatibility tests on different types of UAV platforms and sensors to ensure the stable operation of the architecture. At the same time, optimize the power supply adaptation scheme, dynamically adjust the power distribution according to the power consumption requirements of different hardware, and extend the flight time of the UAV.

This system also sets a unified detection status. For different interface peripherals, the status is reported uniformly as “normal”, “warning”, “fault”, etc., increasing the adaptability of different interface peripherals.

#### 4.3.2. Resource allocation optimization

This system is based on the requirements of emergency rescue scenarios and establishes a dynamic resource allocation mechanism. It adjusts the allocation ratios of computing resources, communication resources, and storage resources in real time according to the signal type, processing priority, and hardware resource status. It adopts lightweight signal processing algorithms to reduce the computational load. For example, it quantizes deep learning models to INT8 format to reduce the model size and computational requirements. It designs a resource scheduling algorithm to allocate high-priority signals to high-performance computing cores and low-priority signals to low-power cores, achieving optimal resource allocation.

This system establishes a resource warning mechanism to monitor the usage of computing power, electricity, and storage in real time. When resources are insufficient, it automatically reduces the collection

frequency and processing accuracy of non-core signals to ensure the stable processing of core signals. At the same time, it optimizes the storage strategy, locally caching the pre-processed key data and transmitting the original data to the backend in real time to reduce local storage occupation and improve resource utilization efficiency.

For the data resource allocation strategy, this system performs dynamic resource allocation based on data priority. It unconditionally transmits specific emergency information and discards signals with lower priority to ensure reasonable resource allocation and reuse.

### **4.3.3. Scenario-based adaptation mechanism**

This system designs scenario-specific parameter configuration templates for different emergency rescue scenarios. By adjusting the collection parameters, transmission strategies, and processing strategies, it enables the architecture to be scenario-adapted. For the fire rescue scenario, it enhances the anti-interference ability of infrared signals, improves the accuracy of fire source positioning, reduces the exposure time of visible light signals, and avoids strong light interference. For the earthquake search and rescue scenario: it strengthens the real-time and stability of positioning signals, improves the ability of visible light signals to identify ruins targets, and increases the sensitivity of life sign detection. For the flood rescue scenario, it optimizes the anti-interference ability of signal transmission, enhances the waterproof fog processing capability of image signals, and ensures the perception effect under complex weather conditions.

In addition to scenario-specific parameter configuration, a scene-adaptive learning mechanism is established. Through training the model with historical rescue data, the architecture can automatically identify the scene type based on environmental characteristics and switch the corresponding parameter configuration without manual intervention. At the same time, it supports user-defined scene parameters to meet the personalized needs of special rescue scenarios, enhancing the flexibility and adaptability of the architecture.

### **4.3.4. Multimodal fusion and collaborative decision optimization**

For the fusion level of multimodal signals, achieve full-dimensional integration of data layer, feature layer and decision layer. The data layer completes the standardization and synchronization of signals; the feature layer extracts the core features of different signals, conducts cross-validation and complementary enhancement; the decision layer generates unified rescue decision instructions based on the fused features, supporting autonomous flight, target locking, and rescue material delivery of unmanned aircraft, etc. For instance, by integrating infrared, visible light and positioning signals, it can automatically identify the location of survivors and environmental risks, generate the optimal rescue path, and coordinate with the UAV flight control system to carry out operations.

In the pre-processing layer process shown in **Figure 2** in the previous text, the flowchart is further refined, and the system adds the coordination relationship among data, features and decisions, making the processing efficiency of multimodal signals higher and saving resources

## **5. Conclusion**

This paper starts from the actual needs of emergency rescue scenarios and presents the following core conclusions. It clarifies that the heterogeneity of multi-modal perception signals in emergency rescue drones,

the lack of modular architecture, insufficient hardware compatibility, and weak fusion and decision-making capabilities are the four core issues in current signal processing, which are interrelated and restrict the overall efficiency of the system. Moreover, it constructs a three-level modular signal processing architecture of perception–transmission–pre-processing, clarifies the functional positioning, design logic and coordination mechanism of each level, and realizes the unified collection, coordinated transmission and efficient preprocessing of multi-modal signals. Additionally, it proposes a series of countermeasures such as hardware adaptation, resource optimization, scenario adaptation, and fusion and decision optimization, which solve the problems of heterogeneous signal collaborative processing, hardware compatibility, and resource shortage from the technical level, and enhance the universality, specialization and reliability of the architecture. Further research in the future can be conducted in the following directions:

- (1) Carry out the engineering implementation verification of the architecture, combining specific unmanned aircraft platforms and emergency rescue scenarios, setting up a testing environment to test the actual operational efficiency of the architecture and optimize parameter configuration;
- (2) Explore the deep integration of artificial intelligence technology and the architecture, introducing technologies such as reinforcement learning and federated learning to achieve intelligent optimization of signal processing and autonomous decision-making, and enhancing the system’s adaptive ability;
- (3) Expand the application scenarios of the architecture, applying it to low-altitude economy, smart transportation, environmental monitoring and other fields to enhance the application scope and value of the architecture;
- (4) Strengthen the research on multi-UAV collaboration and air-ground integration, improving the data sharing and collaborative decision-making mechanism, constructing a large-scale emergency rescue perception network, and providing stronger technical support for extreme disaster rescue.

## Disclosure statement

The authors declare no conflict of interest.

## References

- [1] Wu W, Guo J, Tang Y, 2025, How Drones Are Transforming the 72-Hour Golden Rescue Period. *Knowledge Is Power*, 2025(9): 20–23.
- [2] Li M, Ma W, Zhou Y, et al., 2023, A Method for Unmanned Aerial Vehicle Life Search Based on Multi-sensor Fusion. *Journal of Zhengzhou University (Engineering Science Edition)*, 44(2): 61–67.
- [3] Zhang Y, Wang F, 2024, Research Progress on Multimodal Perception and Interaction Technologies for Embodied Intelligent Unmanned Systems. *Acta Automatica Sinica*, 50(2): 289–305.
- [4] Liu H, Chen L, 2023, Urban Underground Space Emergency Rescue UAV Communication and Perception Technology. *Journal of Underground Space and Engineering*, 19(3): 689–698.
- [5] Liu C, Ma R, Peng M, 2023, Integrated Communication and Sensing for Unmanned Aerial Vehicles: Architecture, Technology and Prospects. *Telecommunications Science*, 39(2): 1–9.
- [6] Chen L, Miao Z, Wang X, et al., 2025, Autonomous Aerial Vehicle Technology and its Applications in the Low-Altitude Economy: A Review. *Robotics*, 47(3): 470–496.
- [7] Han S, Wang M, Duan J, et al., 2024, Research on Unmanned Aerial Vehicle Emergency Communication System

Based on Edge Computing. *Computer Science and Exploration*.

- [8] Wang J, Zhang P, 2023, Relay Technology for Unmanned Aerial Vehicle Emergency Communication Based on Network Coding. *Journal of Communications*, 44(5): 78–90.
- [9] Chen M, Zhao X, 2024, Research on Modular Design and Hardware Adaptation of Multi-modal Perception System for Unmanned Aerial Vehicles. *Computer Engineering and Applications*, 60(4): 123–132.
- [10] Wu F, Sun M, 2022, Compatibility Design of Signal Processing Architecture for Heterogeneous Unmanned Aerial Vehicle Platforms. *Acta Aeronautica et Astronautica Sinica*, 43(S1): 189–198.
- [11] Huang T, Li J, 2023, Lightweight Signal Processing Algorithm and Resource Optimization for Emergency Rescue Unmanned Aerial Vehicles. *Computer Application Research*, 40(7): 2012–2017.
- [12] Ruza J, Savchenko A, Ruza I, et al., 2026, Analysis of World Experience and Experimental Implementation of Unmanned Radio Intelligence Systems. *Transport and Telecommunication Journal*, 27(1): 70–85.
- [13] Ghaly R, Soliman A, Rihan M, 2024, Software-Defined Radio Deployments in UAV-Driven Applications: A Comprehensive Review. *IEEE Open Journal of Vehicular Technology*, 2024(5): 1589–1613.
- [14] Rossi D, Borghi G, Vezzani R, 2025, TakuNet: An Energy-Efficient CNN for Real-Time Inference on Embedded UAV Systems in Emergency Response Scenarios, *Proceedings of 2025 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops*, 339–348.
- [15] Zhen Y, Cheng L, Sun H, et al., 2024, Review on Detection and Recognition of Unmanned Aerial Vehicles Based on Deep Learning. *Signal Processing*, 40(4): 609–624.

**Publisher's note**

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.